Annals of the Rheumatic Diseases 1994; **53**: 789–797

REVIEW

Genetic aspects of familial osteoarthritis

Sergio A Jimenez, Rita M Dharmavaram

Human osteoarthritis (OA) is a heterogeneous and multifactorial disease characterised by the progressive deterioration of the cartilage of diarthrodial joints. Multiple aetiological and pathogenetic mechanisms have been implicated in its development and progression.[1] In many instances OA is an acquired process secondary to various metabolic, mechanical, or inflammatory-immunological events. However, it has long been recognised that several distinct forms are inherited as dominant traits with a Mendelian pattern.[2 3] The most common form of inherited OA is characterised by the presence of Heberden's and Bouchard's nodes and the concentric or uniform degeneration of the articular cartilage of several joints, particularly the hips and knees.[4] Many studies have examined the genetic factors that may be associated with either development or severity of this form of OA. Analysis of the frequencies of HLA antigens in various populations of patients affected have yielded conflicting results,[5–9] although in one study an increased frequency of the HLA-A1B8 phenotype was observed.[9] This study also examined the frequency of α1-antitrypsin phenotypes in these individuals and found a significant increase in the MZ phenotype,[9] although these results are unlikely to be related to the primary genetic defect in this form of OA. A second type of inherited OA is familial chondrocalcinosis—a disease in which calcium pyrophosphate dihydrate (CPPD) crystals are deposited in fibrous and hyaline cartilage.[10] The observations that the degenerative arthritis occasionally precedes or is not associated with demonstrable deposition of CPPD crystals[11] and that there is a physical association of CPPD crystals with components of cartilage matrix[12 13] have led to the suggestion that cartilage matrix abnormalities may be a primary common event leading to cartilage degeneration, CPPD crystal deposition, or both. A third familial form of OA is known as Stickler syndrome or hereditary arthro-ophthalmopathy.[14] This syndrome is characterised by ophthalmological manifestations such as progressive vitreoretinal degeneration and severe myopia, peculiar facial features, and premature degenerative joint disease. Other heritable disorders accompanied by premature OA include hydroxyapatite deposition disease[15] and certain forms of multiple epiphysial dysplasias.[16]

The pattern of inheritance of these diseases is consistent with the hypothesis that mutations in one or more of the genes encoding the macromolecular components of articular cartilage may be responsible for the premature and generalised degeneration of the tissue matrix. The abnormal genes could include the genes for cartilage matrix macromolecules, for enzymes involved in the biosynthesis of matrix, for hormone and growth factor receptors in chondrocytes, or for enzymes involved in the metabolic degradation of the tissue. Recent evidence, however, suggests that the genes encoding the collagenous components of cartilage matrix are the most likely candidates. The collagens represent the most abundant protein of articular cartilage matrix, comprising about 50% of the dry weight of the tissue. These molecules play a crucial role in the maintenance of the biomechanical properties of cartilage, being responsible for the tensile strength and shear stiffness of the tissue. The remarkable complexity of the organisation of the collagenous components in articular cartilage has recently become apparent[17 18] and at least five different collagen types representing the products encoded by at least 10 distinct genes have been recognised in the tissue. Although the cartilage-specific type II collagen is the most abundant collagen species in the tissue, it is very likely that the other collagen types (types VI, IX, X, and XI) also play important structural and functional roles. The normal supramolecular assembly of the various cartilage collagens serves as a mechanical constraint to limit the expansion of proteoglycans and their tightly bound water molecules into the large hydrodynamic domains characteristic of proteoglycans in free solution. A failure of this collagenous assembly would result in swelling of the proteoglycans, increased tissue water, softening of the matrix and eventual degeneration of the cartilage. There is, therefore, compelling evidence to suggest that cartilage collagen genes may be those at fault in heritable OA. Several studies have identified mutations in the genes encoding type I and type III procollagens in patients with other heritable disorders that cause mechanical failure of the connective tissues, such as osteogenesis imperfecta.[19] These investigations have provided a number of successful strategies for the identification of structural mutations in procollagen genes. However, as discussed in more detail below, many of these strategies have limitations when applied to the study of mutations in the cartilage collagen genes, such as the limited availability of cartilage, the difficulty in expanding chondrocyte populations in vitro, the loss of chondrocyte-specific phenotype during culture, and the difficulty in performing

Jefferson Medical College, Thomas Jefferson University, Philadelphia, PA 19107, USA
S A Jimenez
R M Dharmavaram

Correspondence to:
Dr S A Jimenez,
Bluemle Life Sciences Building,
233 South 10th Street,
Rm 509,
Philadelphia,
PA 19107-5541, USA.

extensive protein characterisation from small tissue samples. The following sections will review some of the strategies that have been successfully utilised or that may be potentially useful to identify collagen gene mutations in inherited diseases affecting articular cartilage.

Restriction fragment length polymorphism (RFLP) analysis

The development of recombinant probes that detect polymorphic sites in human DNA by RFLP analysis has made available a vast resource of genetic markers to follow the inheritance of specific DNA sequences in families.[20-22] These polymorphic sequences occur frequently in the flanking regions of most genes, as well as randomly throughout genomic DNA; their detection has made it possible to identify abnormal alleles of many genes and to trace their pattern of co-segregation with a given disease phenotype in families. The premise for RFLP study of inherited forms of OA is that, despite the heterogeneity and variability of their phenotype, they are caused by a primary genetic defect that resides at a specific chromosomal locus. The disease phenotype and the defective gene must, therefore, map to the same chromosomal location as reflected by their co-segregation in families. Using appropriate restriction enzymes it becomes feasible to map the inherited trait to a region in the chromosome bracketed by two specific markers.

The potential benefits of utilising RFLP genetic linkage analysis in heritable forms of human OA is that it will permit the identification of the genes at fault in these disorders. Even the finding of linked genetic markers at some distance from the responsible gene would allow definition of the approximate chromosomal location of the abnormal gene and its cloning and sequencing. In addition, informative RFLPs can be tested in families suffering from diseases with similar phenotypes to determine if the same genetic defect is present in all cases or if there is genetic heterogeneity among the various phenotypes. RFLP analysis can also give valuable negative information, as the absence of co-segregation (i.e. recombination) with the disease phenotype can exclude the candidate gene as the site of the defect in a given family.

Recently, several polymorphisms in the genes for some of the collagens expressed in articular cartilage and in surrounding DNA sequences have been identified. For example, polymorphisms in the genes for type II procollagen (COL2A1), the α2-chain of type IX collagen (COL9A2), and type X collagen (COL10A1) have been described.[23-27] The identification of these polymorphisms has permitted the application of RFLP analysis to test the possibility that structural mutations in various collagen genes expressed in cartilage may be responsible for the biomechanical failure of articular cartilage in familial OA. For example, Francomano *et al*[28] and Knowlton *et al*[29] demonstrated co-inheritance of certain polymorphic sites in COL2A1 with the expression of hereditary arthro-ophthalmopathy or Stickler syndrome. In the study by Knowlton *et al*,[29] three large Stickler syndrome families were analysed for co-inheritance of the clinical manifestations with the *Hind*-III and the variable number tandem repeat (VNTR) polymorphisms in COL2A1. Genetic linkage between the disease phenotype and COL2A1 was demonstrated in the largest family. The results from the second family also supported linkage to COL2A1, leading to the conclusion that mutations in the COL2A1 gene are responsible for the disease in these two families. In conrast, in the third family, recombination between clinical expression and COL2A1 was demonstrated suggesting, therefore, that the syndrome may be heterogeneous and that in certain families a gene other than COL2A1 may be the defective gene.

Subsequently, Knowlton *et al*[30] demonstrated co-inheritance of a phenotype of premature OA and a mild chondrodystrophy with polymorphisms in the type II procollagen gene. Similar results were obtained in another family with primary generalised OA.[31] However, a large study of 61 patients with primary generalised OA failed to show a significant difference in the frequencies of one or more COL2A1 alleles between affected and control individuals.[32] These results thus excluded the possibility that there is a common mutation at COL2A1 that is responsible for the disease phenotype. In a more recent study, Weaver *et al*[33] analysed the inheritance pattern of *Hind*-III and *Hinf*-I RFLPs of COL2A1 in a large five generation family with multiple epiphysial dysplasia (MED). The data indicated that the segregation patterns of COL2A1 and MED phenotype were discordant, with recombination observed in a minimum of two meioses in the portion of the family tested. From these results, it was determined that the MED locus in this family does not map within 2 centimorgans (cM) of COL2A1, therefore excluding this gene as the site of the mutation responsible for the disease phenotype. RFLPs in the type VI collagen gene were also tested for linkage with MED in this family. Recombination between COL6A2 and COL6A3 genes and MED was observed, also ruling them out as candidate genes. Moreover, the data excluded MED from a range of 5 cM on either side of the RFLP marker for COL6A1, effectively excluding COL6A1 as the candidate gene.

Despite the enormous potential of RFLP analysis for the identification of the defective genes in inherited diseases, the method has certain limitations. For example, RFLP analysis depends on the ability to identify large families in whom the disorder is clearly inherited, and from whom blood or tissue samples can be obtained from affected and non-affected members of three or more generations. Also, RFLP analysis does not allow the identification of the exact mutation responsible for the disease phenotype.

Recently, several polymerase chain reaction (PCR) based techniques have been used to type DNA polymorphisms. One method relies

on the fact that in the human genome there are 50 000–100 000 interspersed $(CA)_n$ blocks with n being 15–30. The function of such dinucleotide repeats or microsatellites is unknown, but it is known that the number of repeats within DNA varies among the individuals within a species.[34] It was subsequently shown that blocks of human $(CA)_n$ repeats exhibit length polymorphisms and that these length polymorphisms could be utilised to follow allele co-segregation occurring during meiosis.[35] Subsequently, other short tandem repeats such as tri- and tetranucleotide repeats, although less abundant than $(CA)_n$ markers, have provided an additional source of highly informative markers.[36] The procedure for typing the $(CA)_n$ block markers and other simple sequence repeats is now widely utilised, mainly because of its speed—essentially it involves only two steps: amplification of a DNA fragment by PCR in the presence of radioactive nucleotide and elecrophoresis on a denaturing polyacrylamide gel.

In the MED family studied by Weaver *et al*,[33] described above, in whom COL2A1 and COL6A1 were excluded as candidate genes, such dinucleotide repeat and simple sequence repeat polymorphisms have been utilised for mapping of the disease locus to the pericentromeric region of chromosome 19.[37] In another study, of nine multigenerational families with pseudoachondroplasia, a number of cartilage specific and non-cartilagenous extracellular matrix candidate genes including aggrecan, proteoglycan link protein, small cartilage matrix protein and various collagen genes have been excluded by RFLP analysis. However, pseudoachondroplasia in these families has been mapped to chromosome 19 utilising chromosome 19 microsatellite markers on that chromosome.[38]

Another PCR based method that is being utilised as a screening method for detection of DNA sequence changes is the single strand conformation polymorphism (SSCP) analysis. The method involves amplification by PCR of a discrete segment of genomic DNA in the presence of radiolabelled nucleotides, temperature denaturation of the PCR products, and analysis of single strands on a non-denaturing polyacrylamide gel. Polymorphic differences in strand mobility result from the effects of primary sequence changes on the folded structure of a single DNA strand. The primary sequence differences alter the intramolecular interactions that generate a three dimensional folded structure. The molecules may thus move at different rates through a non-denaturing polyacrylamide gel. Since the conformational variations are subtle, the success of any particular SSCP analysis experiment depends on several factors including the sequence of the DNA fragment being evaluated and optimisation of the experimental conditions to maximise differential migration among fragments. Under optimal conditions the efficiency of SSCP to detect DNA sequence differences such as polymorphisms or mutations has been reported to be close to 100%.[39]

We have previously reported the amplification of the entire coding region of the type X collagen gene, using PCR and SSCP analysis based on the previously reported sequence.[26] We identified a Gly to Arg substitution of the carboxyl (C-) terminal of the protein that results in a loss of a restriction enzyme site. Restriction analysis of a large MED family ruled out this sequence change as a cause of disease. Using a similar approach, Sweetman *et al*[27] have investigated if mutations in COL10A1 are responsible for diverse heritable cartilage diseases including achondroplasia, hypochondroplasia, pseudoachondroplasia and thanatotrophic dysplasia. They identified seven sequence changes in the coding and flanking regions of COL10A1 in affected individuals. However, six of these changes did not co-segregate with the expression of these diseases and were, therefore, polymorphisms rather than true mutations. The sequence changes were also used to demonstrate discordant segregation between the COL10A1 gene and achondroplasia and pseudoachondroplasia. A seventh sequence change that results in a Val to Met substitution at the C-terminal was only found in two individuals with hypochondroplasia from the same family. Segregation analysis of the family was inconclusive because of uncertainty in the diagnosis of one of the family members.

Recently, SSCP analysis followed by genomic DNA sequencing has been used for identification of two mutations in COL10A1 in families with autosomal dominant Schmid metaphysial chondrodysplasia. One of them is a 13 bp deletion of 1857–1869 bp and the other is a T→C change at position 1771 bp that results in a Cys→Arg substitution at amino acid 591 of the protein.[40 41]

Another PCR based method that has been successfully utilised for the identification of mutations in the cartilage collagen genes is known as 'heteroduplex analysis'. This procedure allows detection of mismatches between double stranded DNA containing one wild type strand and a complementary strand from a mutant gene. For this analysis, amplified PCR products of a test sample and wild type control are mixed, heated to denature the double stranded DNA and allowed to reanneal at a lower temperature. If the primary sequences of the sample and control differ, 50% of the reannealed double stranded DNA will be heteroduplex DNA. When separated on a non-denaturing polyacrylamide gel, the reannealed products of the PCR reactions migrate in a distinct pattern: homoduplex strands of mutant and control DNA migrate as a function of their length; heteroduplexes formed of one mutant strand and one wild type strand migrate at a different rate through the gel because the region of mismatch forms a 'kink' in the DNA. Thus a heteroduplex frequently appears on the gel as a distinct band, separate from the homoduplex DNA. Several factors affect the resolution of single base mismatches in heteroduplex DNA, including DNA size (200–600 bp being optimal), position of mismatch within the DNA

fragment (central position being most easily detected), and type and context of mismatch. The advantage of heteroduplex analysis is that it is technically simple, does not require complex chemical or temperature gradients, and bands can generally be detected by ethidium bromide staining. Utilising this method we have recently reported a mutation that results in a 10 bp deletion from 1867–1876 bp in a five generation family with Schmid metaphysial chondrodysplasia.[42]

Identification of type II collagen gene mutations in affected individuals by genomic sequencing

Determination of the sequence of the type II collagen gene has proved successful in certain forms of heritable OA. For example, sequencing exons 2–52 of the gene identified a single base mutation that results in the substitution of Cys for Arg at position 519 of the α1(II) procollagen chain in affected members of a family with primary generalised OA associated with mild chondrodysplasia.[43] A mutation that introduced a stop codon in COL2A1 corresponding to position 732 of the α1(II) procollagen was found in affected members of a family with Stickler syndrome,[44] and two different mutations in exon 48 of COL2A1 have been found in two cases of spondyloepiphysial dysplasia.[45 46] Furthermore, a COL2A1 mutation that results in a Gly to Ser substitution at residue 943 in the C-terminal domain of the triple helix has been identified in a case of achondrogenesis/hypochondrogenesis, a lethal perinatal form of short-limbed dwarfism.[47] The number of newly identified mutations in COL2A1 responsible for a variety of heritable diseases affecting cartilaginous structures has increased rapidly as the methods for gene sequencing have improved in accuracy and speed. The spectrum of mutations in COL2A1 that had been identified by late 1993 are listed in the table.

Mutations in the type II procollagen gene identified in heritable diseases of cartilage (as reported by late 1993)

Disease	*Mutation*	*Reference*
SED	Deletion AA 964 to 999 (exon 48)	45
Achondrogenesis II/ hypochondrogenesis	Gly^{943} → Ser (exon 46)	46
SED	Duplication AA 970 to 984 (exon 48)	47
OA with mild chondrodysplasia	Arg^{519} → Cys (exon 31)	43
SED	†Gly^{997} → Ser (exon 48)	49
Stickler syndrome	Arg^{732} → Stop (exon 40)	44
Hypochondrogenesis	Gly^{574} → Ser (exon 33)	50
Stickler syndrome	Arg^{9} → Stop (exon 7)	51
SED	†Arg^{789} → Cys (exon 41)	52
SED	Arg^{75} → Cys (exon 11)	53
SED	Gly^{247} → Ser (exon 19)	Ritvaniemi *et al* (pers. comm.)
SEMD	Gly^{154} → Arg (exon 15)	54
Stickler syndrome	Frame shift (exon 40)	55
SED	G + 5 IVS20	56
SED	Gly^{493} → Ser (exon 30)	Katzenstein *et al* (pers. comm.)
Hypochondrogenesis	Gly^{583} → Glu (exon 43)	57
Kniest/Stickler syndrome	Deletion IVS12	58
ED/POA	Gly^{976} → Ser (exon 48)	Williams *et al* (pers. comm.)
Hypochondrogenesis	Gly^{988} → Arg (exon 48)	Ganguly *et al* (pers. comm.)
Achondrogenesis II/ hypochondrogenesis	Gly^{691} → Arg (exon 38)	Williams *et al* (pers. comm.)

All these mutations were identified by genomic DNA sequencing except for † identified by sequencing of 'illegitimate' transcripts of the gene. SED = spondyloepiphysial dysplasia; SEMD = spondyloepimetaphysial dysplasia; ED/POA = Epiphysial dysplasia and precocious OA.

Identification of type II procollagen gene mutations by analysis of the corresponding complementary DNA (cDNA)

Each of the mutations described above was identified by sequencing genomic DNA from affected individuals. However, sequencing the large genes encoding for the cartilage specific collagens is time consuming, or is not possible because of the incomplete knowledge of their structure, therefore identification of mutations by sequencing the corresponding cDNAs has many practical advantages. This approach has been very successful and has resulted in the identification of mutations in the genes encoding for the structural connective tissue components in a large number of families with various forms of osteogenesis imperfecta or the Ehlers-Danlos syndrome, and in some families with vascular aneurysms.[19] The approach adopted in most of these instances has been the determination of the nucleotide sequences of cDNAs prepared by reverse transcription of total cellular RNA from cultured cells (fibroblasts or amniotic cells) from affected individuals. Many of these studies were greatly facilitated by the previous identification and characterisation of protein abnormalities which pointed to the likely nature of the mutations and to the approximate regions within the gene containing the mutations. These protein abnormalities included the demonstration of overmodification of post-translational reactions, 'protein suicide' as a result of decreased procollagen thermal stability, the finding of disulphide bonded α-chain dimers resulting from the presence of cysteine residues in the triple helical domains, or the demonstration of shortened collagen α-chains.

Although this method allowed the identification of a number of mutations affecting types I and III collagen genes, its success depends on the availability of large numbers of cells expressing the transcripts of interest, in order to obtain sufficient amounts of RNA for the abnormal protein, and on the information obtained from structural studies of the mutated protein. The availability of cells that express the transcripts from the mutated genes is not a problem for investigation of mutations affecting types I and III collagen genes, because cultured dermal fibroblasts or amniotic cells can be successfully utilised. However, in the case of mutations in articular cartilage collagen genes, this approach has been obstructed by the difficulty in obtaining sufficient tissue to isolate RNA and the loss of cartilage phenotype when chondrocytes are cultured on plastic substrata in vitro.[59]

Amplification and sequencing of human type II collagen cDNA from freshly isolated and from cultured human fetal and adult chondrocytes

The discovery of a specific DNA polymerase that is capable of reiterative synthesis of DNA copies from a given template at high temperatures has resulted in the development

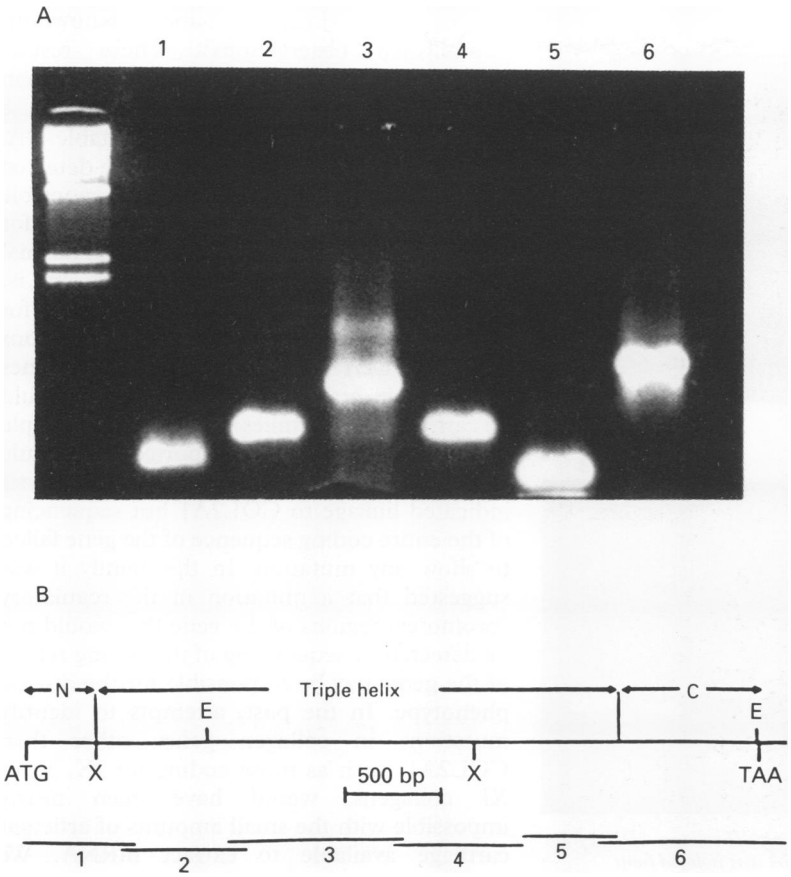

Figure 1 Amplification of the entire type II collagen cDNA from human fetal chondrocytes. cDNA was amplified as described in the text and 10% of the PCR product was loaded on a 1% (w/v) agarose gel and stained with ethidium bromide. Lambda DNA digested with Hind-*III was used as a marker. A: Amplification of cDNA was carried out using the following primers:* LANE 1 *II-1/II-2;* LANE 2 *II-3/II-4;* LANE 3 *II-5/II-6;* LANE 4 *II-7/II-8;* LANE 5 *II-9/II-10;* LANE 6 *II-11/II-12. B: Map of type II collagen cDNA indicating the length and position of the fragments shown in A. bp = Base pairs.*

of one of the most powerful techniques of molecular biology.[60] The PCR reaction has been successfully applied to cloning and sequencing of cDNAs and genes that were not amenable to the techniques previously, because of either the development of secondary structures or the lack of available enzyme restriction sites for proper cloning. The possibilities of identifying cartilage collagen gene mutations using small amounts of cartilage or few cultured chondrocytes have been improved dramatically by the PCR. The amplification of specific DNA sequences can overcome the limitations introduced by the lack of sufficient amounts of articular cartilage necessary to obtain RNA for generation of cDNA libraries. In this approach, minute amounts of RNA can be reverse transcribed to obtain cDNAs, and these cDNAs can then be PCR amplified using sequence specific oligonucleotides based on the known sequence of full length cDNAs. Since the entire sequence of the type II collagen cDNA has been published, it has become feasible to amplify the entire coding sequence from the cDNAs. The PCR products can be sequenced directly or can be cloned into bacterial vectors for standard sequencing procedures. Given the advances in direct DNA sequencing, large stretches of DNA sequence can now be obtained from a single sequencing reaction. Thus to sequence the entire coding region for the mutant allele is feasible; single base mutations or small insertions/deletions in coding or regulatory sequences can be identified; mutations in sequences important for splicing can also be detected utilising oligonucleotide primers specific for intronic sequences at the intron-exon boundaries. Thus the PCR amplified products will correspond to the coding sequences and to adjacent consensus sequences required for splicing.

We have successfully used this approach and have amplified the entire type II collagen cDNA, a length of 4·4 kb, in only six fragments as shown in figure 1, in which the length and position of the PCR fragments shown in A are illustrated in B. Sequencing of these PCR fragments indicated that their sequence was identical to that of corresponding regions of the human type II collagen cDNA reported previously.[61] The RNA used in this experiment was obtained from freshly isolated 24 week old fetal chondrocytes. In order to determine the minimum amount of RNA required to amplify the entire type II collagen cDNA, we performed PCR amplification of RNA in concentrations ranging from 1 μg to 0.1 ng. We found that 1 ng of RNA was the smallest amount required to obtain amplification of each PCR fragment. In this experiment, 1 μg of RNA was obtained from 3×10^5 freshly isolated human fetal chondrocytes; therefore RNA equivalent to that obtained from only 300 chondrocytes was sufficient to obtain amplification of one fragment, or 1800 chondrocytes for amplification of the six fragments encompassing the entire type II collagen cDNA. Amplification of human type II collagen cDNA was also obtained from chondrocytes from a small postmortem specimen of articular cartilage from a 75 year old woman immediately after isolation (fig 2A), and after culture on porous hydroxyapatite ceramic crystals for 90 days (fig 2B). Thus even the small amounts of articular cartilage obtained from arthroplasty or from arthroscopic surgery can be successfully utilised to identify mutations in COL2A1, by use of PCR.

Identification of mutations that do not affect the coding sequence of COL2A1

In addition to the benefits of RFLP genetic linkage analysis to identify the genes that may be at fault in heritable OA, RFLP analysis can give valuable negative information because the absence of co-segregation (i.e. recombination) with the disease phenotype can exclude a candidate gene in a given family. Knowlton *et al*[29] showed the presence of recombination in one of the families with Stickler syndrome they studied. This observation conclusively excluded a mutation in COL2A1 as the cause of the disease phenotype in this family. Similar analysis of another family with Stickler syndrome and of a large family with severe premature OA associated with CPPD deposition failed to show co-segregation of COL2A1 polymorphisms with the disease

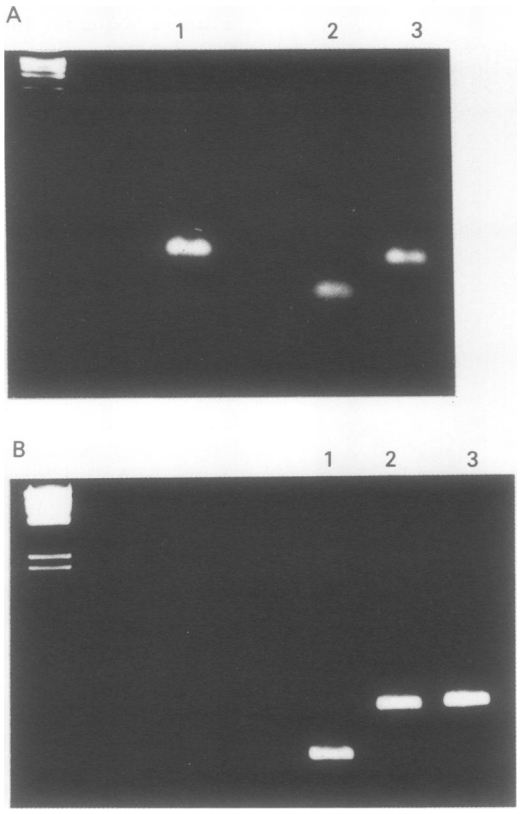

Figure 2 Amplification of the type II collagen cDNA from adult human chondrocytes. RNA was isolated from chondrocytes obtained from knee cartilage of a 75 year old woman. Lambda DNA digested with Hind-*III was used as a marker. A: Amplification of cDNA from fresh chondrocytes was carried out using the following primer pairs: LANE 1 II-3/II-4; LANE 2 II-9/II-10; LANE 3 II-7/II-8. B: Amplification of cDNA obtained from RNA isolated from chondrocytes cultured in porous hydroxyapatite ceramic microcrystals using the following primer sets: LANE 1 II-9/II-10; LANE 2 II-3/II-4; LANE 3 II-7/II-8.*

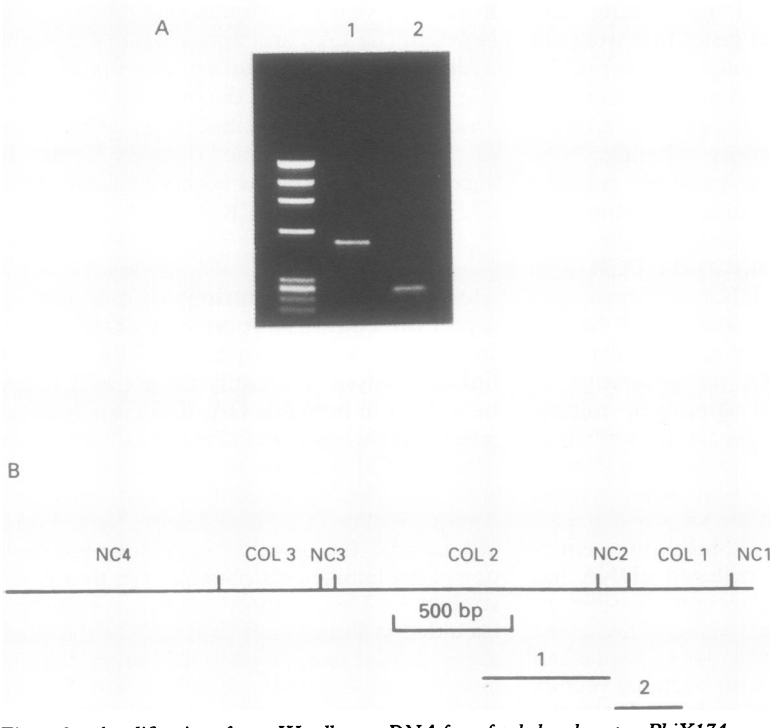

Figure 3 Amplification of type IX collagen cDNA from fetal chondrocytes. PhiX174 DNA digested with Hae-*III was used as a marker. A: The following primer sets were used for amplification: LANE 1 IX-1/IX-2; LANE 2 IX-3/IX-4. B: Map of type IX collagen cDNA indicating the length and position of fragments shown in A. bp = Base pairs.*

phenotype[29] (Jimenez and Knowlton, unpublished observations). These results suggest that mutations in genes other than COL2A1 may be responsible for the disease phenotype in certain subsets of heritable OA. In addition, mutations that cannot be detected by sequencing of the gene may be responsible for the disease. These could involve, for example, mutations that result in abnormal splicing of the mature transcripts. It is, apparent, therefore, that a careful search for mutations that do not affect the coding regions of the COL2A1 or mutations in other genes expressed in articular cartilage matrix should be carried out in families with OA. An example of this has recently been reported by Vikkula *et al.*[62] in a family in whom RFLP analysis indicated linkage to COL2A1 but sequencing of the entire coding sequence of the gene failed to show any mutation. In this family it was suggested that a mutation in the regulatory (promoter) regions of the gene that would not be detected by sequencing of the coding region of the gene may be responsible for the disease phenotype. In the past, attempts to identify mutations in collagen genes other than COL2A1, such as those coding for IX, X, or XI collagens, would have been nearly impossible with the small amounts of articular cartilage available to extract mRNA. We decided, therefore, to examine if the methods developed to amplify cDNA for type II collagen could be successfully utilised to amplify cDNA corresponding to the minor collagens present in the tissue, and attempted to amplify fragments of type IX and type XI collagen cDNA. As shown in figure 3A, we amplified 780 bp of the α1(IX) collagen chain: the length and position of the fragments shown in A are illustrated in B. We also amplified the entire published sequence of the α2(XI) collagen cDNA which corresponds to 67% of the entire cDNA for this chain (fig 4); again, the figure shows the fragments, and their length and position. These studies demonstrated that a combination of PCR with primer-extension/reverse transcription can be utilised to amplify cDNA for the various articular cartilage collagens from mRNA obtained from fresh and cultured human chondrocytes.

From the results shown above, it is apparent that the use of cDNA for identification of mutations in type II and other collagens expressed in articular cartilage has several advantages over genomic sequence analysis. First, the target area to be analysed is much smaller; for example, the entire type II procollagen gene is about 30 kb, whereas the corresponding cDNA is only 4·4 kb. Second, it allows easy detection of mutations in collagen genes the structure of which has not been completely established, such as the genes encoding for types IX, X, and XI collagens. Third, cDNA sequencing allows easy detection of splicing abnormalities such as those which occur when exons are skipped or when aberrant splice sites are used. Finally, the results demonstrate that cDNA analysis of cartilage specific collagens can be successfully

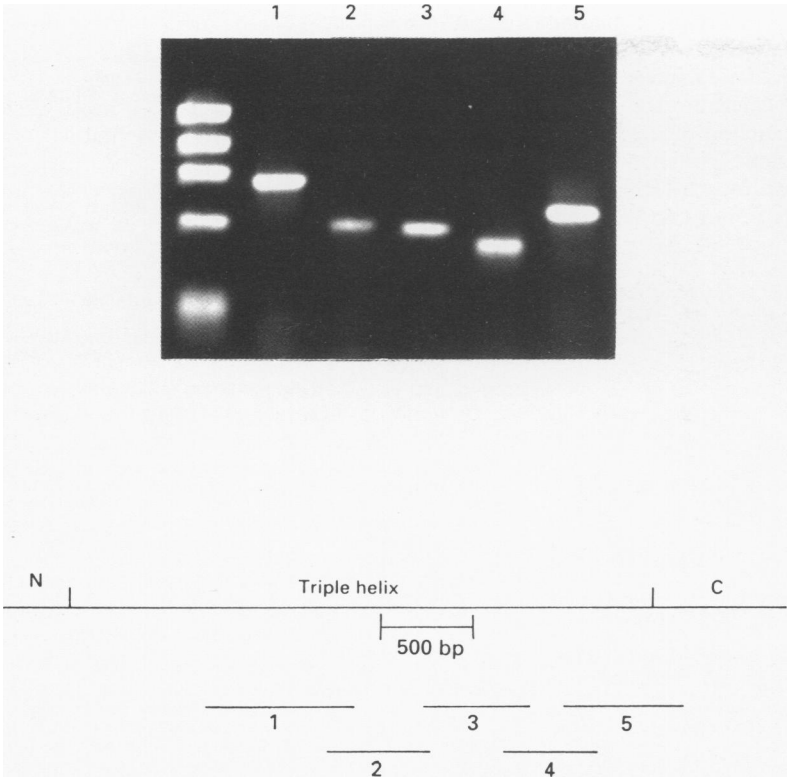

Figure 4 Amplification of type XI collagen cDNA from human fetal chondrocytes. PhiX174 DNA digested with Hae-*III was used as a marker. Amplification of cDNA was carried out using the following primer sets: LANE 1 XI-1/XI-2; LANE 2 XI-3/XI-4; LANE 3 XI-5/XI-6; LANE 4 XI-7/XI-8; LANE 5 XI-9/XI-10. The lower map of type XI collagen cDNA indicates the length and position of the fragments. bp = Base pairs.*

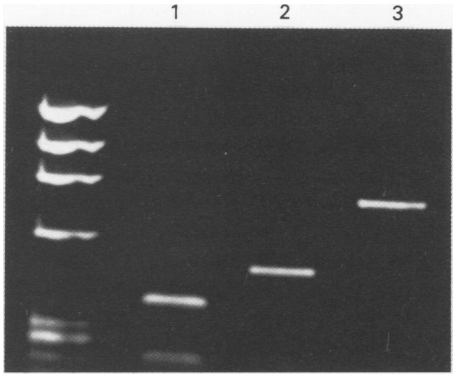

Figure 5 Amplification of type II collagen cDNA from Epstein-Barr virus transformed lymphocytes. PhiX174 DNA digested with Hae-*III was used as a marker. The primer pairs used for amplification were: LANE 1 II-1/II-13; LANE 2 II-9/II-10; LANE 3 II-11/II-14.*

accomplished from small amounts of RNA, as the entire type II cDNA can be amplified from as little as 6 ng of chondrocyte RNA.

New approaches to identify type II collagen gene mutations when articular cartilage tissue is not available

Despite the extraordinary sensitivity of the PCR, the method requires the availability of at least minute amounts of cartilage, or of cultured chondrocytes that maintain their differentiated phenotype. Furthermore, the success of this approach depends on the ability to generate cDNAs that cover the entire coding sequence, or on obtaining shorter cDNAs that cover the region where the mutation is located. Because these problems can often be difficult to overcome, approaches that do not depend on the availability of cartilage may be required in the majority of cases. One of these approaches was used recently to identify a mutation that results in a Gly to Ser substitution at residue 997 of the α1(II) procollagen chain in a patient with spondyloepiphysial dysplasia congenita, using PCR amplification of the corresponding cDNA from 'illegitimate transcripts' obtained from fibroblasts and Epstein-Barr virus (EBV)-transformed lymphocytes.[49] However, from these studies it was not clear if the transcripts contained the full length type II collagen mRNA, as only 621 bp of the cDNA were sequenced. We recently used this approach to amplify cartilage specific cDNAs from mRNA obtained from non-cartilaginous sources.[26] As shown in figure 5, we obtained amplification of cDNA fragments corresponding to the amino-terminal (lane 1), the triple helical (lane 2), and the C-terminal (lane 3) domains of the type II collagen cDNA from RNA extracted from EBV-transformed lymphocytes. Sequencing of 1575 bp indicated that the illegitimate transcripts were correctly spliced. We also demonstrated, for the first time, illegitimate transcription of the genes encoding for types IX and XI collagens in EBV-transformed lymphocytes, as we were able to amplify cDNA fragments corresponding to the C-termini of both of these collagens. However, we were unable to obtain any amplification of cartilage specific collagen cDNAs from RNA obtained either from untransformed lymphocytes or from lymphocytes induced to proliferate by exposure to phytohaemagglutinin (results not shown). Because of the very small amounts of correctly spliced mRNA in lymphoblasts (less than 1 mRNA molecule per 500-1000 cells), large amounts of starting material (at least 500 ng RNA) and increased number of PCR cycles were required for amplification of cDNA from low abundance transcripts such as those for types II, IX, and XI collagens. The increased number of PCR cycles could increase the error in amplified DNA sequences, as Taq DNA polymerase has a relatively high rate of single base misincorporation. However, misincorporated bases can be easily detected by sequencing at least five separate clones from each of two independent amplifications. Furthermore, this problem can be overcome by the use of direct sequencing of PCR products with which the authentic sequence is in vast excess over the sequences with misincorporated bases.

Conclusion

The explosive advances in the application of molecular biological methods have allowed the identification of a large number of gene mutations in various heritable diseases. The well known occurrence of certain subsets of OA displaying an autosomal pattern of inheritance has rendered these diseases amenable to similar studies.

The methods reviewed here have already allowed the identification of a large number of

mutations in the genes encoding articular cartilage collagens in patients with various inherited disorders of articular cartilage (listed in the table) and will permit the rapid characterisation and sequencing of cDNA for the various cartilage specific collagens from small amounts of chondrocyte RNA, or from RNA obtained from non-cartilaginous sources.

The expert assistance of R Tate and O Ma in the preparation of the manuscript is gratefully acknowledged. Supported by NIH Program Project Grant PO1-AR39740.

1 Moskowitz R W. Clinical and laboratory findings in osteoarthritis. In: McCarty D J, Koopman W J, eds. *Arthritis and allied conditions*, 12th ed. Philadelphia: Lea & Febiger, 1993: 1735–60.
2 Stecher R M, Hersh A H, Hauser H. Heberden's nodes. The family history and radiographic appearance of large family. *Am J Hum Genet* 1953; **5:** 46–69.
3 Kellgren J H, Lawrence J S, Bier F. Genetic factors in generalized osteoarthritis. *Ann Rheum Dis* 1963; **22:** 237–55.
4 Marks J S, Stewart I M, Hardinge K. Primary osteoarthritis of the hip and Heberden's nodes. *Ann Rheum Dis* 1979; **38:** 107–11.
5 Lawrence J S, Gelsthorpe K, Morell G. Heberden's nodes and HLA markers in generalized osteoarthritis. *J Rheumatol* 1983; **10** (suppl 9): 32–3.
6 Brodsky T, Appelboam A, Govaerts Famey J P. Antigens HLA et nodosites d'Heberden. *Acta Rheumatologica* 1979; **3:** 95–105.
7 Ercilla M G, Brancos M A, Breysse Y, *et al.* HLA antigens in Forestier's disease, ankylosing spondylitis, and polyarthrosis of the hands. *J Rheumatol* 1977; **4** (suppl 3): 89–93.
8 Benavides G, Cervantes A, Silva B, Katona G, Lardizabal J. HLA and Heberden's nodes in Mexican mestizos. *Clin Rheumatol* 1985; **4:** 97–8.
9 Pattrick M, Manhire A, Ward A M, Doherty M. HLA-A, B antigens and α1-antitrypsin phenotypes in nodal generalized osteoarthritis and erosive osteoarthritis. *Ann Rheum Dis* 1989; **48:** 470–5.
10 Ryan L M, McCarty D J. Calcium pyrophosphate crystal deposition disease: pseudo-gout; articular chondrocalcinosis. In: McCarty D J, Koopman W J, eds. *Arthritis and allied conditions*, 12th ed. Philadelphia: Lea & Febiger, 1993: 1835–55.
11 Bjelle A O. Morphological study of articular cartilage in pyrophosphate arthropathy. *Ann Rheum Dis* 1972; **31:** 449–56.
12 Bjelle A O. Cartilage matrix in hereditary pyrophosphate arthropathy. *J Rheumatol* 1981; **8:** 959–64.
13 Beutler A, Rothfuss S, Clayburne G, Sieck M, Schumacher H R. Calcium pyrophosphate dihydrate crystal deposition in synovium. Relationship to collagen fibers and chondrometaplasia. *Arthritis Rheum* 1993; **36:** 704–15.
14 Stickler G B, Belau P G, Farrell F J, *et al.* Hereditary progressive arthro-ophthalmopathy. *Mayo Clin Proc* 1965; **40:** 433–55.
15 Marcos J C, DeBenyacar M A, Garcia-Morteo O, Maldonado-Cocco J A, Morales V H, Laguena R P. Idiopathic familial chondrocalcinosis due to apatite crystal deposition. *Am J Med* 1981; **71:** 557–64.
16 Spranger J. The epiphyseal dysplasias. *Clin Orthop Rel Res* 1975; **114:** 46–59.
17 Mayne R. Cartilage collagens. *Arthritis Rheum* 1989; **32:** 241–6.
18 Eyre D R. Structure and function of cartilage collagens. In: Brandt K D, ed. *Cartilage changes in osteoarthritis.* Indianapolis: Indiana University Press, 1990: 12–6.
19 Prockop D J. Mutations that alter the primary structure of type I collagen. *J Biol Chem* 1990; **265:** 15349–52.
20 Botstein D, White R L, Skolnick M, Davis R W. Construction of a genetic linkage map in man using restriction fragment length polymorphisms. *Am J Hum Genet* 1980; **32:** 314–31.
21 Willard H F, Skolnick M H, Pearson P L, Mandel J-L. Report on the committee on human gene mapping by recombinant DNA techniques. *Cytogenet Cell Genet* 1985; **40:** 360–489.
22 Antonarakis S. Diagnosis of genetic disorders at the DNA level. *N Engl J Med* 1989; **320:** 153–63.
23 Sangiorgi F O, Benson-Chanda V, de Wet W J, Sobel M E, Tsipouras P, Ramirez F. Isolation and partial characterization of the entire human pro-α1(II) collagen gene. *Nucleic Acids Res* 1985; **13:** 2207–25.
24 Väisänen P, Elima K, Pelotie A, Peltonen L, Vuorio E. Polymorphic restriction sites of type II collagen gene: their location and frequencies in the Finnish population. *Hum Hered* 1988; **38:** 65–71.
25 Weaver E J, Knowlton R G. A PvuII polymorphism near the 5' end of the type II procollagen gene (COL2A1). *Nucleic Acids Res* 1989; **17:** 6429.
26 Dhamavaram R M, Peng M, Strawbridge R R, Jimenez S A. Polymerase chain reaction amplification of the coding sequences of the type X collagen gene from genomic DNA identification of a polymorphism that changes Gly to Arg at position 545 by single-strand conformation polymorphism analysis. *Biochem Biophys Res Comm* 1992; **187:** 420–4.
27 Sweetman W A, Rash B, Sykes B, *et al.* SSCP and segregation analysis of the human type X collagen gene (COL10A1) in heritable forms of chondrodysplasia. *Am J Hum Genet* 1992; **51:** 841–9.
28 Francomano C A, Liberfarb R M, Hirose T, *et al.* The Stickler syndrome: evidence for close linkage to the structural gene for type II collagen. *Genomics* 1987; **1:** 293–6.
29 Knowlton R G, Weaver E J, Struyk A F, *et al.* Genetic linkage analysis of hereditary arthro-ophthalmopathy (Stickler syndrome) and the type II procollagen gene. *Am J Hum Genet* 1989; **45:** 681–8.
30 Knowlton R G, Katzenstein P L, Moskowitz R W, *et al.* Demonstration of genetic linkage of the type II procollagen gene (COL2A1) to primary osteoarthritis associated with a mild chondrodysplasia. *N Engl J Med* 1990; **322:** 526–30.
31 Palotie A, Vaisanen P, Ott J, *et al.* Predisposition to familial osteoarthrosis linked to type II collagen gene. *Lancet* 1989; **1:** 924–7.
32 Priestley L, Fergusson C, Ogilvie D, *et al.* A limited association of generalized osteoarthritis with alleles at the type II collagen locus: COL2A1. *Br J Rheum* 1991; **30:** 272–5.
33 Weaver E J, Summerville G P, Yeh G, *et al.* Exclusion of type II and type VI procollagen gene mutations in a five generation family with multiple epiphyseal dysplasia. *Am J Med Genet* 1993; **45:** 345–52.
34 Tyler-Smith C, Brown W R A. Structure of the major block of alphoid satellite DNA on the human Y chromosome. *J Mol Biol* 1987; **195:** 457–70.
35 Weber J L, May P E. Abundant class of DNA polymorphisms which can be typed using the polymerase chain reaction. *Am J Hum Genet* 1989; **44:** 388–96.
36 Edwards A, Civitello A, Hammond H A, Caskey C T. DNA typing and genetic mapping with trimeric and tetrameric tandem repeats. *Am J Hum Genet* 1991; **49:** 746–56.
37 Oehlmann R, Summerville G P, Yeh G, Weaver E J, Jimenez S A, Knowlton R G. Genetic linkage mapping of multiple epiphyseal dysplasia to the pericentromeric region of chromosome 19. *Am J Hum Genet* 1994; **54:** 3–10.
38 Hecht J T, Francomano C A, Briggs M D, *et al.* Linkage of typical pseudoachondroplasia to chromosome 19. *Genomics* 1993; **18:** 661–6.
39 Sheffield V C, Beck J S, Kwitek A E, Sandstrom D W, Stone E M. The sensitivity of single-strand conformation polymorphism analysis for the detection of single base substitutions. *Genomics* 1993; **16:** 325–32.
40 Warman M L, Abbot M, Apte S S, *et al.* A type X collagen mutation causes Schmid metaphyseal chondrodysplasia. *Nature Genet* 1993; **4:** 79–82.
41 McIntosh I, Hefferon T W, Warman M L, Olsen B R, Francomano C A. Mutation of type X collagen (COL10A1) resulting in Schmid metaphyseal chondrodysplasia. *Am J Hum Genet* 1993; **52** (suppl): A1201.
42 Dharmavaram R M, Elberson M A, Peng M, Kirson L A, Kelly T E, Jimenez S A. Identification of a mutation in type X collagen in a family with Schmid metaphyseal chondrodysplasia. *Hum Mol Genet* 1994; **3:** 507–9.
43 Ala-Kokko L, Baldwin C T, Moskowitz R W, Prockop D J. Single base mutation in the type II procollagen gene (COL2A1) as a cause of primary osteoarthritis associated with a mild chondrodysplasia. *Proc Natl Acad Sci USA* 1990; **87:** 6565–8.
44 Ahmad N N, Ala-Kokko L, Knowlton R G, *et al.* Stop codon in the procollagen II gene (COL2A1) in a family with the Stickler syndrome (arthro-ophtalmopathy). *Proc Natl Acad Sci USA* 1991; **88:** 6624–7.
45 Lee B, Vissing H, Ramirez F, Rogers D, Rimoin D. Identification of the molecular defect in a family with spondyloepiphyseal dysplasia. *Science* 1989; **244:** 978–80.
46 Tiller G, Rimoin D L, Murray L W, Cohn D H. Tandem duplication within a type II collagen gene (COL2A1) exon in an individual with spondyloepiphyseal dysplasia. *Proc Natl Acad Sci USA* 1990; **87:** 3889–93.
47 Vissing H, D'Alessio M, Lee B, Ramirez F, Godfrey M, Hollister D W. Glycine to serine substitution in the triple helical domain of pro-α1(II) collagen results in a lethal perinatal form of short-limbed dwarfism. *J Biol Chem* 1989; **264:** 18625–7.
48 Pun Y L, Moskowitz R W, Lie S, *et al.* Clinical correlations of osteoarthritis associated with a single-base mutation (Arginine^{519} ›Cysteine) in type II procollagen gene. *Arthritis Rheum* 1994; **37:** 264–9.
49 Chan D, Cole W G. Low basal transcription of genes for tissue-specific collagens by fibroblasts and lymphoblastoid cells. Application to the characterization of a glycine 997 to serine substitution in alpha 1(II) collagen chains of a patient with spondyloepiphyseal dysplasia. *J Biol Chem* 1991; **266:** 12487–94.
50 Horton W A, Machado M A, Ellard J, *et al.* Characterization of a type II collagen gene (COL2A1) mutation identified in cultured chondrocytes from human hypochondrogenesis. *Proc Natl Acad Sci USA* 1992; **89:** 4583–7.
51 Ahmad N N, McDonald-McGinn D M, Zackai E H, *et al.* A second mutation in type II procollagen gene (COL2A1) causing Stickler syndrome (arthro-opthalmopathy) is also a premature termination codon. *Am J Hum Genet* 1993; **52:** 39–45.
52 Chan D, Taylor T K, Cole W A. Characterization of an arginine 789 to cysteine substitution in alpha 1(II) collagen chains of a patient with spondyloepiphyseal dysplasia. *J Biol Chem* 1993; **268:** 15238–45.

53 Williams C J, Considine E L, Knowlton R G, *et al.* Spondyloephyseal dysplasia and precocious osteoarthritis in a family with an $Arg^{75} \rightarrow Cys$ mutation in the procollagen type II gene (COL2A1). *Hum Genet* 1993; **92:** 499–505.
54 Vikkula M, Ritvaniemi P, Vuorio A F, Kaitila I, Ala-Kokko L, Peltonen L. A mutation in the amino terminal end of the triple helix of type II collagen causing severe osteochondrodysplasia. *Genomics* 1993; **16:** 282–5.
55 Brown D M, Nichols B E, Weingeist T A, Kimura A E, Sheffield V C, Stone E M. Type II procollagen gene mutation in Stickler syndrome. *Am J Hum Genet* 1992; **51** (suppl): A301.
56 Tiller G E, Weis M A, Eyre D R, Rimoin D C, Cohn D H. An RNA splicing mutation (G + 5IVS20) in the gene for type II collagen (COL2A1) produces spondyloepiphyseal dysplasia congenita. *Am J Hum Genet* 1992; **51** (suppl): A37.
57 Bogaert R, Tiller G E, Weis M A, *et al.* An amino acid substitution ($Gly^{853} \rightarrow Glu$) in the collagen alpha 1(II) chain produces hypochondrogenesis. *J Biol Chem* 1992; **267:** 22522–6.
58 Winterpacht A, Hilbert M, Schwaze U, Mundles S, Spranger J, Zabel B U. Kniest and Stickler dysplasia phenotypes caused by collagen II gene (COL2A1) defect. *Nature Genet* 1993; **3:** 323–6.
59 Benya P D. Introduction and survey of techniques for chondrocyte culture. In: Maroudas A, Kuettner K, eds. *Methods in cartilage research.* London: Academic Press, 1990: 85–9.
60 Saiki R K, Bugawau T L, Horn G T, Mullis R B, Erlich H A. Analysis of enzymatically amplified beta-globin and HLA-DQ alpha DNA with allele-specific oligonucleotide probes. *Nature* 1986; **324:** 163–6.
61 Baldwin C T, Reginato A M, Smith C, Jimenez S A, Prockop D J. Structure of cDNA clones coding for human type II procollagen. The $\alpha 1$(II) chain is more similar to the $\alpha 1$(I) chain than two other α chains of fibrillar collagens. *Biochem J* 1989; **262:** 521–8.
62 Vikkula M, Palotie A, Ritvaniemi P, *et al.* Early-onset osteoarthritis linked to the type II procollagen gene. *Arthritis Rheum* 1993; **36:** 401–9.